Proficiency Estimation by Motion Variability obtained from Single Camera Input

Ayumi Matsumoto, Dan Mikami, Harumi Kawamura, and Akira Kojima
NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Kanagawa, Japan

Keywords: Motor Learning, Proficiency Estimation, Variability of Movements, Markerless Motion Capture.

Abstract: This paper proposes a motor learning assist system that estimates the proficiency of trainees in making sports motions on the basis of variability of his/her own 3D motions in trials captured by a single camera. Most existing systems assume that a sequence of human poses must be obtained by multiple calibrated cameras or a marker-based motion capture system. Such systems can be effectively used by professional athletes or broadcast station personnel who specialize in sports, but not by casual sports fans who have no particular athletic skills. We propose a method for evaluating proficiency on the basis of obtained for trainees in repeated trials. Usable by not only elite athletes but also casual sports fans, the method has two important features. First, it requires only a freely positionable single camera since its 3D pose estimation methodology is independent of camera position. Second, it estimates proficiency only from a trainee's own motions and thus does not require any reference movements. In this paper, the golf swing is used as the target of motor learning. Experiment results show that variability in 3D motion in trials is inversely proportional to the test subjects' degree of proficiency.

1 INTRODUCTION

Computer vision-based sports assistance has been widely studied in recent years (Yu and Farin, 2005). Reported techniques, include player tracking (Pingali et al., 1998), semantic annotation of sports game videos (Assfalg et al., 2002), and sports event detection (Luo et al., 2003). In addition, a real-time tracking technique called Hawk-eye (Owens et al., 2003) is now being used for assisting umpires in making judgements.

Computer vision assistance in motor learning for casual sports fans, amateure sports teams, and artists has also been considered. Chua et al. proposed a system for learning sports motions by using virtual reality (Chua et al., 2003). Choi et al. proposed a system that compares the motions of a trainee with those of an expert to provide information that is useful for practicing sports (Choi et al., 2008). So far, however, these systems have rarely been used in practice.

One main reason such systems have not gained popularity is they require making troublesome preparations. Most existing systems assume that a sequence of human poses must be obtained by multiple calibrated cameras or a marker-based motion capture system. Such systems can be effectively used by professional athletes or broadcast station personnel who specialize in sports, but not by casual sports fans who have no particular athletic skills.

Aiming to overcome the need for troublesome preparations, Mikami et al. proposed a method for analyzing human motions from a video sequence captured by a single camera (Mikami et al., 2013; Mikami et al., 2012). Their method enables analysis of subtle changes in repetitive motions, but to apply it the relative positions of the camera and the subject should be the same when capturing motions. To address these problems, we propose a system that can provide assistance in motor learning to all sports fans without troublesome preparations.

Generally speaking, it can be assumed that motor learning is divided into three phases: cognitive, associative, and autonomous (Schmidt and Lee, 1988). In the cognitive phase, a trainee seeks to make correct movements through a trial-and-error process. In the associative phase, a trainee knows the check points to follow to make correct movements and how to modify incorrect movements. In the autonomous phase, the motions become ingrained; the trainee can make them almost automatically without thinking about them.

On the basis of this assumption, we propose a method for evaluating proficiency based on the vari-

Matsumoto A., Mikami D., Kawamura H. and Kojima A..
Proficiency Estimation by Motion Variability obtained from Single Camera Input.
DOI: 10.5220/0004608301470153
In *Proceedings of the International Congress on Sports Science Research and Technology Support* (icSPORTS-2013), pages 147-153
ISBN: 978-989-8565-79-2

ability of the trial when it is subjected to repeated trials. When the motor learning is in the cognitive phase, we assume there is large variation in movements. When the learning progresses and moves to the associative phase, we assume the variation becomes rather small because a trainee gets to know the check points and how to modify incorrect movements. Finally, when it reaches the autonomous phase, we assume that the variation in movements becomes quite small.

To achieve our purpose, we use a markerless motion capture system proposed by Matsumoto et al. (Matsumoto et al., 2012) that can obtain human poses in 3D world coordinates from one camera input that is robust under camera positions. This feature enables motion variations to be compared from different views for the same criteria. In this paper, we use the golf swing as the motor learning target since there are so many persons who enjoy the game of golf.

The reminder of this paper is organized as follows. Section 2 overviews the markerless human motion capture method proposed by Matsumoto et al., Section 3 outlines our proficiency estimation method in detail, Section 4 describes experiments we conducted, and Section 5 concludes the paper with a brief summary.

2 GPDM-BASED MARKERLESS MOTION CAPTURE WITH SINGLE CAMERA

Our method uses a GPDM-based markerless motion capture system proposed as a method for obtaining human poses under robust camera position (Matsumoto et al., 2012). Consisting of a training step and a pose estimation step, it estimates the poses made when subject movements are similar to trained movements.

The method we propose has two principal features. First, it requires only a single camera to work well. Due to the innate characteristics of single camera input, it is unable to obtain depth information. To compensate for this, it includes a training step in which use is made of 3D motion data captured by a marker-based motion capture system. 3D motion data of various motions in sports can be found in databases that are publicly available. For example, the CMU mocap library includes motions of golf swings, basketball dribbles, and soccer ball kicks. Our method can employ 3D motion data from such motion databases and thus skirt troublesome preparations. Second, it is robust against differences in the relative positions of the camera and the target human. The GPDM-based markerless motion capture method learns the state dynamics of all possible views. In the pose estimation step, it jointly estimates a 3D human pose and the camera position relative to it. This enables robust tracking against camera position. The following subsections describe each of these steps in more detail.

2.1 Training Step

The GPDM-based markerless motion capture method uses sequences of 3D human poses in the training step. That is, it requires pose sequences obtained by a marker-based motion capture system or a multi-camera system. Note that since our system is designed for specific motor learning, this requirement is not especially problematic.

The training step actually comprises three steps. The first is estimating a view-dependent trajectory. It should be noted that the a view-dependent observation can be virtually generated from training data because it consists of 3D motion data. The second is reducing the dimensionality of data by using GPDM (Wang et al., 2005). The third is learning the state dynamics for each view. Through this training step, the view-dependent dynamics of human movement in a low-dimensional feature space are obtained.

2.2 Pose Estimation Step

In the pose estimation step, state parameters, i.e., view and pose parameters trained in the training step, are estimated by particle filtering (Isard and Blake, 1998) of a video sequence captured by a single camera. Because this step estimates not only pose parameters but also the view (i.e., the relative positions of human and camera), it is robust with respect to the latter. An HSV histogram of joints was used for the observation model.

3 PROPOSED METHOD

This section outlines the details of the proficiency estimation system we propose for assisting motor learning. It has been said that motor learning can be divided into three phases: cognitive, associative, and autonomous (Schmidt and Lee, 1988). On the basis of this knowledge, the proposed method estimates a trainee's proficiency in making motions from the variability of his/her own movements.

Figure 1 shows the system environment that we assume. The proposed method uses a single cam-

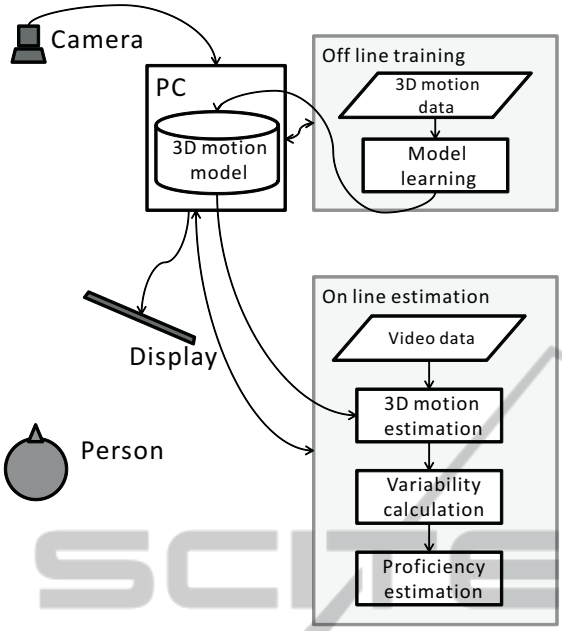


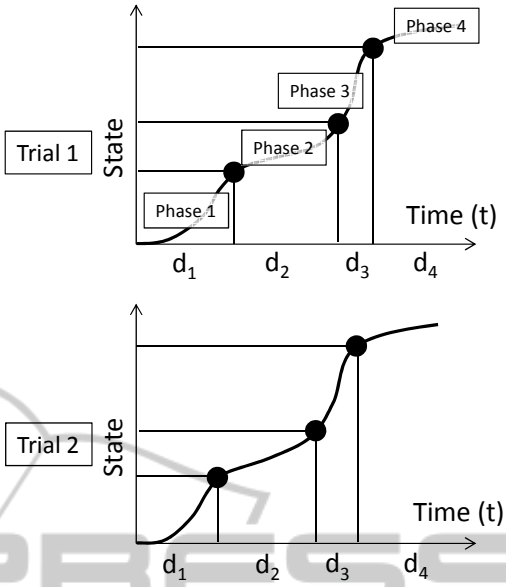
Figure 1: Assumed system environment. The proposed method captures a trainee's movements with a single camera. The video is transferred to a PC, which then estimates 3D human poses and the view on the basis of a pre-training 3D motion model. Through multiple pose estimation trials, the variability of movements is calculated and then proficiency is estimated and displayed. This simple setting is a most important feature in practical use.

era to capture a trainee's movements. The video is transferred to a PC, which then uses the method described in Sect. 2 to estimate 3D human poses and the view on the basis of a pre-training 3D motion model. Through multiple pose estimation trials, the variability of movements is calculated and then proficiency is estimated and displayed.

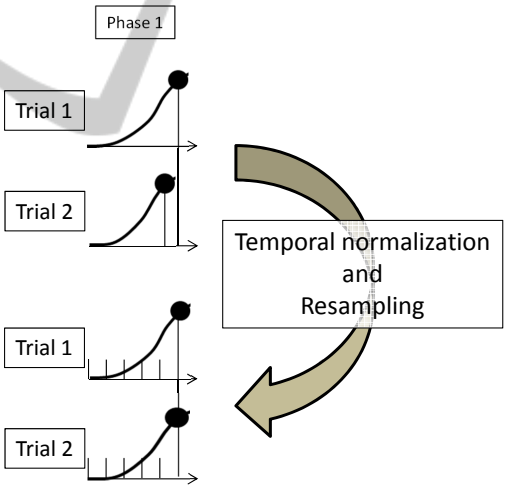
The variability estimation step consists of a phase division step and a variability calculation step. The proposed method first divides the movements into predefined phases. The variability calculation step then calculates variabilitices of duration, trajectory, and pose at the switching point of phases. Specifically, for a motion that has N phases, the method yields $3N+1$ variabilities: N for duration, N for trajectory, and $N+1$ for poses of terminal points. Figure 2 illustrates the proposed variability estimation.

3.1 Phase Division Step

Phase division is motion dependent and requires knowledge of the motions being made. In this subsection, we use the swinging of a golf club and the throwing of a ball to show examples of phase division. Note that the phase definitions given are merely



(a) For each trial, a motion is divided into phases; in this figure, the two trials (Trial 1 and Trial 2) displayed are divided into four phases. The variability estimation step calculates variabilities of duration, trajectory, and pose at the switching point of the phases.



(b) Duration varies by trial. Therefore, to calculate the variability of trajectory, the proposed method coordinates the durations of trials. It then resamples by a specified time step.

Figure 2: Variability estimation of our method.

examples.

3.1.1 Golf Swing

As shown in Fig. 3, a golf swing consists of three phases: backswing, downswing, and follow-through. The backswing starts with the picture labeled "setup" and ends with the one labeled "to". The downswing goes from "top" to "impac" and the follow-through

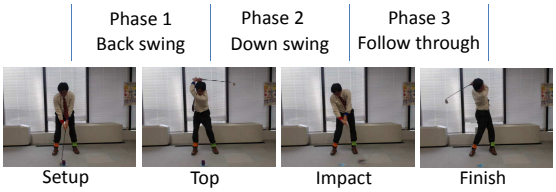


Figure 3: Three phases of golf swing; back swing, down swing, and follow-through.

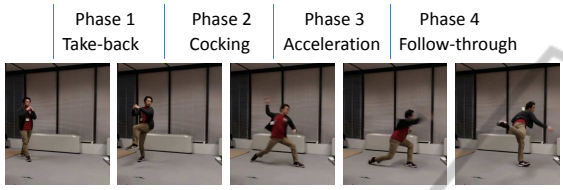


Figure 4: Four phases of ball throwing: take-back, cocking, acceleration, and follow-through.

from “impact” to “finish”.

3.1.2 Ball throwing

As shown in Fig. 4, the action of throwing a ball consists of four phases: take-back, cocking, acceleration, and follow-through.

These phase definitions are associated with body configuration parameters. Therefore, by using the sequence of pose estimates, automatic phase division can be executed.

3.2 Variability Calculation Step

The variability calculation step calculates variabilities of duration $V(d)$, trajectory $V(t)$, and pose at the switching point of phases $V(p)$. Because the motions are divided into phases in the phase division step, variabilities of duration $V(d)$ and pose $V(p)$ can be calculated by normal form.

For calculating the variability of trajectory $V(t)$, the proposed method uses a three-step solution. First, it coordinates the durations of trials. Second, it resamples by a specified time step s. If a resampled time does not have an observed pose data, it estimates the pose by interpolation. Finally, it calculates the average of variabilities at resampled time steps.

4 EXPERIMENTAL

To verify the effectiveness of the proposed method, we conducted experiments in which the golf swing was used as the target motion. In this section we first show the experimental settings and then the results.

4.1 Experimental Settings

4.1.1 Subjects

In the experiments, each subject was asked to swing a golf club ten times. The subjects varied in their experience in playing golf as shown in Table 1. Since the main topic of this study was proficiency estimation, we asked the subjects to wear colored bands on their wrists and ankles to facilitate stable tracking.

4.1.2 Motion Data and Capturing

We used golf swing data of an adult male of average build from the CMU motion capture library (Subject #64) in the training step. The frame rate of this motion data was 120 fps.

We captured videos of the subjects’ motions with two cameras at different angles with 640 x 480 pixel resolution. Capturing frame rate was 30 fps and each trial had about 60 frames.

We downsampled the motion data to eliminate the gap between its frame rate and that for the captured videos.

4.1.3 Calculation

We used an Intel Core i7 3.20GHz CPU for calculation. For all experiments, 3000 particles were used for the particle filtering to estimate 3D human pose. The phases were automatically divided on the basis of the pose estimates; the phase lengths were normalized in 5 samples. We caluculated the variability from ten trials in each phase.

4.2 Result of 3D Pose Estimate

Snapshots of captured videos are shown in Fig. 5. The overlay lines denote pose estimates. Note that the estimates included 3D pose information and the view; the lines were generated from 3D poses and projected on the basis of the estimated view. We verified that we successfully obtained correct 3D poses despite the difference of view point. Figure 6 (the solid lines) shows left hand horizontal positions in 3D pose coordinate. The different markers on the lines denote camera location. The estimated positions were almost same results in despite of a different camera position.

We also calculated 2D color tracking results to show the benefit of using 3D poses in Fig. 6 (the dotted lines). The position provided at a different camera place was different in the timing of the peak. This is

Table 1: Subject s golf experiences.

Subjects	Experience years	Frequency	Average score
A	18	Once a week	95
B	7	Five times a year	118
C	0	None	None

Figure 5: Results of 3D pose estimation shown by red lines. Top: Camera 1, Subject C. Bottom: Camera 2, Subject A.

because that information is degenerated nonlinearly by obtaining 3D position with a camera. This result affects subsequent calculation, for example, the division for a phase. This means that using 3D poses allows an equal trajectory to be obtained independent of the camera position and that the proposed method is not affected by the camera position relative to the trainee; this is an important feature for practical use.

4.3 Result of Proficiency estimation on the Basis of Motion Variability

Figure 7 shows the estimated variability of motions. The variability was calculated from results obtained in seven trials (out of ten, after removal of unstable tracking). As Fig. 7 shows, the well experienced subject A showed very small variability of motions, while the beginner subject C showed large variability. The tendency that dispersion became small so that an acquisition degree became higher was provided.

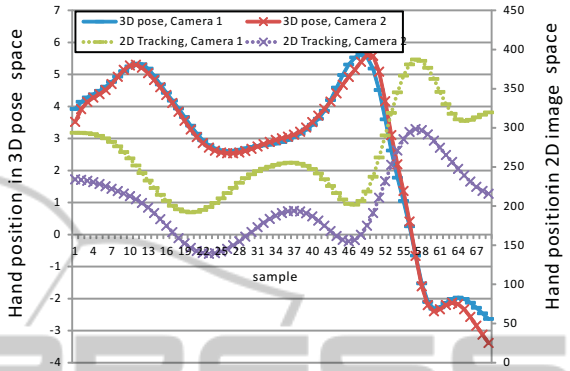
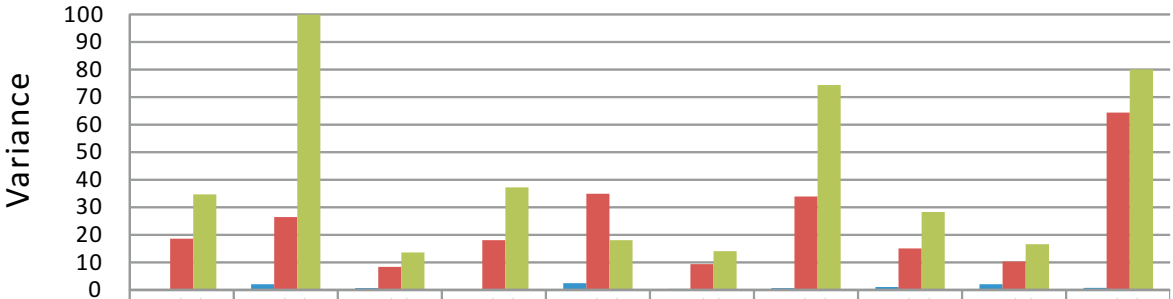


Figure 6: Results of left hand horizontal position calculated from 3D estimated pose in 3D pose coordinate and 2D color tracking results in image coordinate. Solid lines denote 3D estimated pose, dotted lines denote 2D color tracking, and different markers on lines denote camera location. The position provided at a different camera place was the same by 3D pose evaluation substantially, but the timing of the peak is different in the 2D tracking.

4.4 Discussion

In these experiments, the elapsed time for proficiency calculation from ten trials was about fifty minutes. Since we will attempt to use this system to obtain real-time feedback for assisting motor learning, we assume we need to obtain about a 50-fold reduction in the computational time. This would appear to be quite difficult but we believe it is not unachievable. Particle filters are very suitable for use in parallel computing techniques. In recent research there have been many reports of using GP-GPU to achieve a more than 10-fold increase in particle filter computation speed.

To enable the proposed method to estimate proficiency with greater precision, we consider it needs to be able to deal with two important types of information that it currently does not. One is the variability of motions within a phase. The method currently calculates the variability of trajectory on the basis of the average of pose variabilities within a phase. However, the variability of motions within a phase may also be informative. An example illustrating this point is shown in Fig. 8. The results show that the well experienced subject A showed small variability at all times, while subjects B and C showed significant vari-



	V(p) Set-up	V(d) Back-swing	V(t) Back-swing	V(p) Top	V(d) Down-swing	V(t) Down-swing	V(p) Impact	V(d) Follow-through	V(t) Follow-through	V(p) Finish
Subject A	0.33	2.12	0.69	0.14	2.49	0.44	0.70	1.06	2.07	0.81
Subject B	18.61	26.48	8.41	18.12	34.94	9.37	33.96	15.11	10.33	64.37
Subject C	34.67	107.14	13.63	37.25	18.08	14.10	74.38	28.35	16.66	80.07

Figure 7: Estimated variability of motions. $V(d)$ denotes variabilities of duration, $V(t)$ denotes the average of variability of trajectory at each time step, and $V(p)$ denotes the average of variability of pose at each time step. Experienced subject showed small variability of motions, while the beginner subject showed large variability.

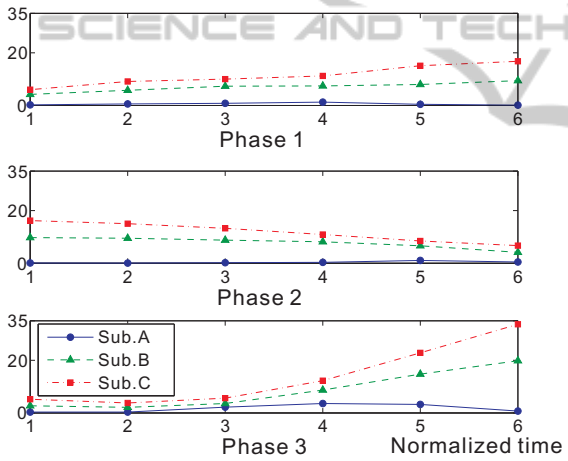


Figure 8: Variability of motions within an experiment phase. Results show that the well experienced subject A showed small variability at all times, while subjects B and C showed significant variability with a similar tendency.

ability with a similar tendency. We only show these results at this time, but more detailed results may be obtained by using a more precise analysis method in the acquisition process. The other is coordination between body parts. It may be considered that the motion of one part is changed to compensate for that of another part, so that the total performance is adjusted well. Therefore, the method's effectiveness could be enhanced by enabling it to handle this kind of coordination between body parts.

5 CONCLUSIONS

We have proposed a motor learning assist system that estimates the proficiency of trainees in performing motions such as those made in swinging a golf club or throwing a ball. In this paper, we proposed a method for evaluating proficiency on the basis of variability in 3D motion obtained for trainees in repeated trials, under the assumption that less variability over a number of trials correlates with the trainees' proficiency. Its principal features are that first, it requires only a single camera and can be used not only by elite athletes but also by casual sports fans, and second, it estimates proficiency only from a trainee's own motions and thus does not require any reference movements. We found there was a tendency for dispersion over 10 trials to fall in proportion to the increased degree of proficiency the trainees acquired.

In future work we will need to face two challenges. One is to validate our assumption that small/large variability leads to good/poor performance. The other is to enhance our system's effectiveness. We will examine means to accelerate its calculation speed, perform experiments on it with an increased number of subjects including experts, and confirm its applicability to other types of motions. In addition, we want to inspect vision-feedback methods to be effective in the improvement of the athletic ability.

REFERENCES

Assfalg, J., Bertini, M., Colombo, C., and Bimbo, A. D. (2002). Semantic annotation of sports videos. *IEEE Multimedia*, 9(2):52–60.

Choi, W., Mukaida, S., Sekiguchi, H., and Hachimura, K. (2008). Quantitative analysis of Iaido proficiency by using motion data. In *ICPR*, pages 1–4.

Chua, P. T., Crivella, R., Daly, B., Hu, N., Schaaf, R., Ventura, D., Camill, T., Hodgins, J., and Pausch, R. (2003). Training for physical tasks in virtual environments: Tai Chi. In *IEEE Virtual Reality 2003*, pages 87–94.

Isard, M. and Blake, A. (1998). CONDENSATIN-conditional densitiy propagation for visual tracking. *IJCV*, 29(1):5–29.

Luo, Y., Wu, T.-D., and Hwang, J.-N. (2003). Object-based analysis and interpretation of human motion in sports video sequences by dynamic bayesian networks. *CVIU*, 92(1-2):196–216.

Matsumoto, A., Wu, X., Kawamura, H., and Kojima, A. (2012). 3D motion estimation of human body from video with dynamic camera work. In *MPRSS*, pages 71–78.

Mikami, D., Kimura, T., Kadota, K., Kashino, M., and Kashino, K. (2012). Inter-trial difference analysis through appearance-based motion tracking. In *30th International Society of Biomechanics in Sports*.

Mikami, D., Kimura, T., Kadota, K., Kawamura, H., and Kojima, A. (2013). Human motion analysis under actual sports game situations -sequential multi-decay motion history image matching-. In *VISAPP 2013*.

Owens, N., Harris, C., and Stennett, C. (2003). Hawk-eye tennis system. In *Internationl Conference on Visual Information Engineering*, pages 182–185.

Pingali, G. S., Jean, Y., and Carlbom, I. (1998). Real time tracking for enhanced tennis broadcasts. In *IEEE CVPR*, pages 260–265.

Schmidt, R. A. and Lee, T. (1988). *Motor control and learning*. Human Kinetics.

Wang, J. M., Fleet, D. J., and Hertzmann, A. (2005). Gaussian process dynamical models. In *NIPS*, pages 1441–1448.

Yu, X. and Farin, D. (2005). Current and emerging topics in sports video processing. In *ICME*, pages 526–529.